

# ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПЕРЕДБАЧЕННЯ ФІШИНГОВИХ САЙТІВ

<sup>1</sup> Вінницький національний технічний університет

## **Анотація**

Розглядається інформаційна технологія передбачення фішингових сайтів за такими ознаками як довжина URL та кількість символів «.», «,», «-», «\_», «+», «\*», що міститься в URL посиланні сайту. Передбачання здійснюється за допомогою методів машинного навчання шляхом вирішення задачі бінарної класифікації.

**Ключові слова:** фішингові сайти, ознаки фішингу, класифікація, інформаційна технологія, кібербезпека, шахрайство.

## **Abstract**

The article considers the information technology for predicting phishing sites based on such features as the length of the URL and the number of characters “.”, “,”, “-”, “\_”, “+”, “\*” contained in the URL link of the site. Prediction is performed using machine learning methods by solving a binary classification problem.

**Keywords:** phishing sites, signs of phishing, cybersecurity, fraud.

## **Вступ**

Фішинг – це форма інтернет-шахрайства, яка стала досить поширеною в останні роки. За цією схемою зловмисники, вдаючись до соціальної інженерії, використовують підроблені веб-сайти та електронні повідомлення, щоб виманити конфіденційну інформацію від потенційних жертв. У цьому контексті важливо розрізняти фішингові атаки від достовірних запитів, щоб уникнути неприємних наслідків для особистої безпеки та фінансового стану.

Для виявлення фішингових сайтів в роботі пропонується використати такі методи машинного навчання як логістична регресія, Decision Tree, Random Forest та Gradient boosting trees.

## **Постановка задачі**

Для виявлення фальшивих веб-сайтів використаємо наступні ознаки [1]:

- url\_length: Довжина URL-адреси.
- n\_dots: Кількість символів '.' в URL-адресі.
- n\_hyphens: Кількість символів '-' в URL.
- n\_underline: Кількість символів '\_' в URL.
- n\_slash: Кількість символів '/' в URL-адресі.
- n\_questionmark: Кількість символів '?' в URL-адресі.
- n\_equal: Кількість символів '=' в URL-адресі.
- n\_at: Кількість символів '@' в URL-адресі.
- n\_and: Кількість символів '&' в URL-адресі.
- n\_exclamation: Кількість символів '!' в URL-адресі.
- n\_space: Кількість символів ' ' в URL-адресі.
- n\_tilde: Кількість символів '~' в URL-адресі.
- n\_comma: Кількість символів ',' в URL-адресі.
- n\_plus: Кількість символів '+' в URL-адресі.
- n\_asterisk: Кількість символів '\*' в URL-адресі.
- n\_hastag: Кількість символів '#' в URL-адресі.
- n\_dollar: Кількість символів '\$' в URL.
- n\_percent: Кількість символів '%' в URL-адресі.
- n\_redirection: Кількість перенаправлень в URL-адресі.
- phishing: Мітки URL-адреси. 1 - фішинговий, 0 - легітимний.

Необхідно за даними ознаками передбачити фішинг сайту. Передбачання зводиться до вирішення задачі бінарної класифікації на основі датасету [2].

### Результати дослідження

Для вирішення задачі бінарної класифікації передбачання фішингу сайту використано такі методи машинного навчання як логістична регресія, Decision Tree, Random Forest та Gradient boosting trees [3]. Результати дослідження наведені в таблиці 1.

Таблиця 1 Результати дослідження

№	Назва методу	Точність класифікації, %
1	Логістична регресія	85
2	Decision Tree	88
3	Random Forests	89
4	Gradient boosting trees	89

З таблиці 1 видно, що метод Gradient boosting trees є найкращим методом передбачання. Точність класифікації на тестовій вибірці складає 89%.

### Висновки

В роботі запропонована інформаційна технологія передбачання фішингу сайту такими методами машинного навчання як логістична регресія, Decision Tree, Random Forest та Gradient boosting trees. Дослідження показали, що метод Gradient boosting trees є найкращим методом.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Як вберегти себе від шахрайства [Електронний ресурс] – Режим доступу до ресурсу: <https://cases.media/en/article/yak-vberegti-sebe-vid-shakhraistva-oznaki-fishingovikh-saitiv>
2. Web Page Phishing Dataset [Електронний ресурс] – Режим доступу до ресурсу: <https://www.kaggle.com/danielfernandon/web-page-phishing-dataset/data?select=web-page-phishing.csv>
3. Штовба С.Д., Козачко О.М. Machine learning: стартовий курс : електронний навчальний посібник. Вінниця : ВНТУ, 2020. – 81 с

*Литвиненко Данило Олександрович* — студент групи 2ІСТ-206, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: lytdanya@gmail.com.

*Козачко Олексій Миколайович* — доц. кафедри системного аналізу та інформаційних технологій. Вінницький національний технічний університет

Науковий керівник: *Козачко Олексій Миколайович* — доц. кафедри системного аналізу та інформаційних технологій. Вінницький національний технічний університет, м. Вінниця

*Lytvynenko Danylo O.* - student of group 2IST-20b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: lytdanya@gmail.com.

*Kozachko Oleksii M.* - Associate Professor, Department of System Analysis and Information Technologies. Vinnytsia National Technical University

Research advisor: *Kozachko Oleksii Mykolaiovych* - Associate Professor, Department of System Analysis and

